MORAL HAZARD IN HEALTH INSURANCE:
WHAT WE KNOW AND HOW WE KNOW IT

Liran Einav
Amy Finkelstein

Moral Hazard in Health Insurance: What We Know and How We Know It
Liran Einav and Amy Finkelstein
NBER Working Paper No. 24055
November 2017
JEL No. D12,G22

ABSTRACT

We describe research on the impact of health insurance on healthcare spending ("moral hazard"), and use this context to illustrate the value of and important complementarities between different empirical approaches. One common approach is to emphasize a credible research design; we review results from two randomized experiments, as well as some quasi-experimental studies. This work has produced compelling evidence that moral hazard in health insurance exists – that is, individuals, on average, consume less healthcare when they are required to pay more for it out of pocket – as well as qualitative evidence about its nature. These studies alone, however, provide little guidance for forecasting healthcare spending under contracts not directly observed in the data. Therefore, a second and complementary approach is to develop an economic model that can be used out of sample. We note that modeling choices can be consequential: different economic models may fit the reduced form but deliver different counterfactual predictions. An additional role of the more descriptive analyses is therefore to provide guidance regarding model choice.

Liran Einav
Stanford University
Department of Economics
579 Serra Mall
Stanford, CA 94305-6072
and NBER
leinav@stanford.edu

Amy Finkelstein
Department of Economics, E52-442
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and NBER
afink@mit.edu

1. Introduction

Empirical work in applied microeconomics is often loosely classified into two categories: "reduced form" or "structural".[1] While this classification is somewhat subjective, surely imperfect, and to some extent artificial – there is a richer spectrum of empirical approaches that could be broken down to many more than two categories – this simple classification is often used to imply two mutually exclusive approaches that are at odds with each other. And the researcher - faced with a question and a data set - is portrayed as needing to make an almost religious choice between the two approaches. In this paper we try to make the simple point – appreciated by many, but perhaps not all – that these two empirical approaches are in fact complements, not substitutes. Each has its own pros and cons. They should often be used in tandem (within or across papers) as scholars embark on answering a specific research question.

To illustrate this point, we use the specific topic of moral hazard in health insurance, on which there is a vast empirical literature (including our own) covering a range of empirical approaches. In the context of health insurance, the term "moral hazard" is widely used (and slightly abused) to capture the notion that insurance coverage, by lowering the marginal cost of care to the individual (often referred to as the out-of-pocket price of care), may increase healthcare use (Pauly 1968). In the United States – the context of all the work we cover in this paper – a typical health insurance contract is annual and concave. It is designed so that the out-of-pocket price declines during the year, as the cumulative use of healthcare increases.

We make no attempt to review the voluminous empirical literature on the topic. Rather, we select only a few specific papers – drawing (grossly) disproportionately on our own work – to illustrate the relationship and complementarities between different empirical approaches used to study the same topic. Our focus is thus not only on describing (some of) what we know, but also on how we know it.

We begin by defining the object of interest: what "moral hazard" means in the context

---

[1]The precise definitions of these two terms is not always clear but it's safe to say that most current empirical micro researchers would agree with Justice Potter Stewart's assessment of hard-core pornography: "I know it when I see it." The reader can judge for herself in the specific applications we discuss below.

of health insurance, and why it is of interest to economists. We then discuss work on three specific questions related to moral hazard in health insurance. First, we describe work that has tested whether moral hazard in health insurance in fact exists. There is a clear affirmative answer, with much of the most-convincing existing evidence coming from large-scale randomized experiments: Just like almost any other good, individuals increase their healthcare utilization when the price they have to pay for it is lower. Second, we describe work that tries to assess the nature of the consumer response. In particular, we ask whether individuals respond to the dynamic incentives that arise from the non-linear health insurance contracts. Again, the general finding is positive, with much of the evidence driven by quasi-experimental studies. Finally, we describe work that attempts to forecast what healthcare spending would be under contracts we do not observe in the data. This requires a more complete model of individual behavior.

In the final section, we conclude by returning to our main goal in writing this paper, and discuss the cross-pollination across the methods and approaches used in the three preceding sections. While all methods were used in the context of the same broad topic, the more specific questions they answer are slightly different. We highlight the value of each approach, and the important interactions between them. In particular, compelling "reduced form" causal estimates of the impact of health insurance contracts on healthcare spending are invaluable for testing specific hypotheses, such as whether there is any behavioral response or whether individuals respond to dynamic incentives. There are settings and questions in which such reduced form estimates may be sufficient. In particular, if the variation used is sufficiently close to prospective policies of interest, one might need to go no further. Yet, many – perhaps most – questions of interest require us to make predictions out of sample, for which economic models that rely on deeper economic primitives are important. These modeling choices should not be made in a vacuum; the descriptive evidence from the reduced form provides general motivation, as well as more specific guidance, as to which modeling choices are more appropriate in a given context.

We are clearly not the first to attempt to highlight the value of combining different empirical approaches in the context of the same question. Very similar views are expressed in Chetty (2009), Heckman (2010), Nevo and Whinston (2010), and Einav and Levin (2010),

among others. While tastes or skill sets of individual researchers may understandably lead them to disproportionately or exclusively pursue one particular style of empirical work, the literature as a whole benefits enormously from attempts to incorporate and cross-pollinate the two, within and across papers. Discussing these issues in the abstract is often difficult, so customizing the discussion to a specific context may be useful. Our modest goal in this paper is to provide such a specific context within which to illustrate this more general point.

2. "Moral Hazard" in Health Insurance

Throughout this paper, we follow decades of health insurance literature and use the term "moral hazard" to refer to the responsiveness of healthcare spending to insurance coverage. The use of the term in this context dates back at least to Arrow (1963). Consistent with the notion of hidden action, which is typically associated with the term "moral hazard," it has been conjectured that health insurance may induce individuals to exert less (unobserved) effort in maintaining their health. For example, Ehrlich and Becker (1972) modeled health insurance as reducing individuals' (unobserved) effort in maintaining their health; because health insurance covers (some of) the financial costs that would be caused by poor health behaviors, individuals may have less incentive to avoid them – they may exercise less, eat more cheeseburgers, and smoke more – when they have insurance coverage.

However, this so-called "ex ante moral hazard" has received very little subsequent attention in empirical work from the literature.[2] This may be because it is not empirically relevant in many contexts – the increased financial cost associated with poor health is not the only cost, and probably not the most important cost of being sick.

The focus of the moral hazard literature has instead been on what is sometimes referred to as "ex post moral hazard." That is, on the responsiveness of consumer demand for healthcare to the price she has to pay for it, conditional on her underlying health status (Pauly 1968;

---

[2]Spenkuch (2012) provides one of the few pieces of evidence on "ex ante moral hazard." He re-analyzes King et al.'s (2009) randomized evaluation of the impact of encouraging individuals in some geographic areas of Mexico but not in others to enroll in the then-newly introduced catastrophic health insurance program for workers outside the formal sector, Seuguro Popular. Spenkuch (2012) finds some evidence of declines in preventive care, such as flu shots and mammograms, associated with experimentally-induced greater insurance coverage.

Cutler and Zeckhauser 2000). In that sense, the use of the term "moral hazard" is a bit of an abuse of the "hidden action" origin of the term. The "action" – i.e., the individual's healthcare utilization – is in fact observed (and contractible), and the asymmetric information problem may be more naturally described as a problem of "hidden information" (regarding the individual's health status). Yet, to stay consistent with decades of abuse of terminology in the entire health insurance literature, we use the term in a similar way and by "moral hazard" refer to how consumer demand for healthcare responds to the out-of-pocket price the consumer has to pay for that care.

Consumer cost-sharing is the typical name used for determining the out-of-pocket price the consumer has to pay for healthcare. Because the set of healthcare services is broad, and the price of each service could vary, insurers often specify coverage as a percentage share of the total healthcare spending. The share of total healthcare spending paid by the individual is referred to as "consumer cost-sharing"; the remaining share is paid by the insurer. For example, a 20% consumer co-insurance or cost-share means that for every dollar of healthcare spending, the consumer pays 20 cents out of pocket and the insurer pays 80 cents.

Typical health insurance contracts are annual and do not specify a constant consumer cost-share. Rather, they specify the consumer cost-sharing as a function of the cumulative (over the covered year) amount of healthcare spending. This function is typically concave. Figure 1 shows a stylized example of a typical contract. This example shows a concave, piece-wise linear schedule with three "arms." In the first – the deductible range – the individual faces an out-of-pocket price of 100%; every dollar of healthcare spending is paid fully out of pocket. After the deductible is exhausted, which in this example occurs at $500 in total spending, the individual enters the "co-insurance" arm, where she faces a price of 10%; for every dollar of healthcare spending. Finally, once the individual has spent a total of $3,500 out of pocket (or $30,500 in total spending), she reaches the "out-of-pocket maximum" (also known as "stop loss" or "catastrophic coverage") arm, at which point she faces no cost-sharing and has complete insurance coverage.

Moral hazard is of economic interest because it creates an obstacle to the consumption-smoothing purpose of insurance. Insurance is valuable because it creates a vehicle for transferring consumption from (contingent) states with low marginal utility of income (e.g. when

4

one is healthy) to states with high marginal utility of income (e.g. when one is sick). The first best insurance contract would equalize marginal utility across different states; the existence of moral hazard makes it infeasible to obtain the first best. As Pauly (1968) first pointed out, if individuals' healthcare utilization responds to the price they have to pay for it and the underlying health status is not contractible, the cost of providing insurance will rise and individuals may no longer be willing to pay the break-even price of full insurance. Therefore, as shown by Holmstrom (1979), the presence of moral hazard leads optimal insurance contracts to be incomplete, striking a balance between reducing risk and maintaining incentives.

A declining out-of-pocket price schedule (see e.g. Figure 1) is a natural way to optimally trade off the goal of combating moral hazard through higher consumer cost-sharing with the goal of providing risk protection through lower consumer-cost sharing. Since the value of insurance is increasing in the total spending, it makes sense to provide a policy that provides greater protection when spending is greater. While this concave feature is common in many health insurance contracts in the United States, we will also discuss below settings where contracts deviates from this pattern.

The existence, magnitude, and nature of the moral hazard response is thus a key input into the optimal design of private or public health insurance contracts. This is a natural reason for the study of moral hazard to attract the considerable theoretical and empirical attention that it has. However, moral hazard in health insurance has also attracted academic and policy interest for the potential it raises that higher consumer cost-sharing could help reduce the high - and rising - levels of healthcare spending as a share of GDP in most developed countries. This has prompted, for example, policy interest in high-deductible health insurance plans in the U.S. as a way of reducing aggregate healthcare spending levels. The majority of healthcare spending, however, is accounted for by a small share of high-cost individuals whose spending is largely in the "catastrophic" range where deductibles and co-payments no longer bind. This suggests that - for meaningful impacts on health care spending - the incentives for health insurance for providers - rather than for consumers - may be more important. The impact of provider incentives in health insurance has, to date, received comparatively less attention than consumer incentives, but there is hope that

this may be changing (see e.g. Clemens and Gottlieb 2014; Eliason et al., 2016; Einav, Finkelstein, and Mahoney 2017). The empirical approaches we discuss here in the context of consumer incentives - and the strong complementarity across them - have natural application to provider incentives.

3. Is There Moral Hazard in Health Insurance?

We now know what moral hazard in health insurance is (or at least what we've all come to call it) and why it could be important for affecting the optimal design of health insurance contracts. But does it exist? Does health insurance actually increase healthcare spending? Health insurance, by design, lowers the price individuals pay for their medical care. First-year economics teaches us that demand curves tend to slope down, that when we make something cheaper, people tend to buy more of it. So the answer may seem obvious. Yet, in the context of healthcare, there are (at least) two views that are less sure.

One view holds that healthcare cannot be analyzed like any other good. Demand for healthcare, in this view, is determined by "needs," not by economic factors, or as an economist might put it, the demand for healthcare is completely inelastic with respect to its price. Malcolm Gladwell has expressed this view forcefully in a New Yorker article tellingly entitled "The Moral Hazard Myth" (Gladwell 2005). Expounding his central premise – that the "myth" of moral hazard in health insurance is a singularly American obsession that has created our singular lack of universal coverage – he writes "The moral hazard argument makes sense ... only if we consume healthcare in the same way that we consume other consumer goods, and to [some] ... this assumption is plainly absurd. We go to the doctor grudgingly, only because we're sick."

There is also a second view, according to which the demand for healthcare in fact slopes up! One version of this conjecture is that health insurance will improve people's health by increasing timely and effective medical care (e.g. preventive care or better management of chronic conditions), and that this improved health will in turn reduce healthcare utilization. Another version points to the efficiency of healthcare use as a channel through which healthcare spending will fall when insurance coverage becomes more generous. For example, while

most healthcare providers in the United States can choose whether or not to see patients, emergency rooms cannot; the Emergency Medical Treatment and Active Labor Act (EMTALA) requires that hospitals provide emergency medical treatment to all patients. There is therefore widespread speculation that one of the benefits of providing health insurance to previously uninsured individuals is to get them out of the expensive emergency room and into cheaper primary care (Dudiak 2013; Palm-Houser 2013; Snyder 2013). Indeed, this idea that insuring the uninsured will reduce expensive (and presumably inefficient or unnecessary) emergency room visits has been a leitmotif of advocates of expanding health insurance coverage in the United States. For example, in making the case that Michigan should expand Medicaid coverage under the Affordable Care Act, Republican Governor Rick Snyder's policy team argued "Today, uninsured citizens often turn to emergency rooms for non-urgent care because they don't have access to primary care doctors – leading to crowded emergency rooms, longer wait times and higher cost. By expanding Medicaid, those without insurance will have access to primary care, lowering costs and improving overall health" (Snyder 2013).

We thus have three widely-circulated competing claims: health insurance increases, decreases, or does not change healthcare spending. Research allows us to move from rhetoric to reality. Ultimately, the existence and sign of any moral hazard effects of health insurance is an empirical question. It's a challenging empirical question because people who have more generous health insurance presumably differ in other ways from people with less generous health insurance, and these differences may be correlated with expected healthcare spending. Indeed, the basic theory of adverse selection suggests that those who have more health insurance are on average in worse health (and hence face higher expected healthcare spending) than those with less health insurance (Akerlof 1970; Rothschild and Stiglitz 1976; Einav and Finkelstein 2011). How to separate such potential selection effects from the treatment effect of interest, namely moral hazard?

We describe evidence from two randomized evaluations of the impact of health insurance on healthcare spending: the RAND Health Insurance Experiment from the 1970s, and the 2008 Oregon Health Insurance Experiment. We review the evidence from each, which shows that moral hazard exists: health insurance increases healthcare spending. We then describe quasi-experimental evidence of moral hazard that uses the existence of "bunching" at a

convex kink in the budget set created by the health insurance contract to also establish the presence of moral hazard (i.e. a behavioral spending response to the contract). We discuss the institutional setting for the RAND Experiment and the "bunching"estimator in some detail, since we will describe further analyses of both these settings in more depth in subsequent sections.

3.1. Two Randomized Evaluations

The Oregon Health Insurance Experiment. In 2008, the state of Oregon engaged in a limited expansion of one of its Medicaid programs. Medicaid is the public health insurance program for low-income individuals in the United States. The particular program in Oregon was available to low-income, uninsured adults, aged 19-64, who were not already eligible for Medicaid by virtue of meeting one of its categorical requirements. This Medicaid program provided comprehensive health insurance coverage with zero consumer cost-sharing. Faced with budgetary constraints that precluded their offering the program to all eligible individuals, policymakers in the state of Oregon decided that a random lottery drawing would be the fairest way to allocate their limited Medicaid slots. The lottery was publicly advertised, and eligible individuals were encouraged to sign up. About 75,000 individuals signed up for the lottery, from which approximately 30,000 were randomly selected. Those who were selected won the ability to apply for Medicaid, and to subsequently enroll in Medicaid if found eligible. About 60% of those selected sent in applications, and about half of those applications were deemed eligible for Medicaid, resulting in about 10,000 individuals who won the lottery and enrolled in Medicaid. The remaining 45,000 who were not selected by the lottery became the control group; they were essentially unable to apply for Medicaid. About two years after the 2008 lottery, the state found additional resources and began to offer the ability to apply to Medicaid to those in the control group.

The lottery created the opportunity to use a randomized controlled design to study the effects of Medicaid coverage over its first two years. Specifically, random assignment by the lottery can be used as an instrument for Medicaid coverage (Imbens and Angrist 1994). Over the approximately two-year study period, lottery assignment increased the probability of having health insurance coverage by about 25 percentage points. Using this experimentally-

induced variation in insurance coverage, researchers have studied the short-term effects of Medicaid on a wide range of outcomes. The evidence indicates that Medicaid increases healthcare spending, improves economic security, and improves some health measures. We focus here on a subset of the healthcare spending results [3]

The results from the experiment show that Medicaid increases healthcare spending across the board, including hospital admissions, emergency department visits, primary care, preventive care, and prescription drugs. Illustrating a subset of these findings, Figure 2 shows the increased use of the emergency department (top panel) and the increase in primary and preventive care (bottom panel). Both panels plot the mean of the control group against that mean plus the "local average treatment effect" estimate of Medicaid, i.e. the estimate of the impact of Medicaid on the outcome, using winning the lottery as an instrument for Medicaid coverage. For example, the estimates indicate that Medicaid increases the probability of having a primary care visit in the last 6 months by 21 percentage points, or over 35% relative to the control group's mean, and the probability of having a recommended mammogram in the last 12 months by 19 percentage points, or about 65%. A back-of-the-envelope calculation using the induced increases in healthcare utilization suggests that, in the first year, Medicaid increases annual healthcare spending by about $775, or about 25% per year (Finkelstein et al. 2012).

The effect appears to operate across all types of care, with estimated increase in both "high value" care (such as preventive care) as well as in potentially "low value" care (such as emergency room visits for non-emergency conditions).[4] Indeed, contrary to the argument that Medicaid would *decrease* emergency department visits, the evidence indicates that Medicaid in fact *increased* emergency department visits by 40%; this increase in emergency department visits occurs across all kinds of patients (e.g. those who had used the emergency room frequently prior to the experiment and those who had not recently been) and all kinds of visits (e.g. on-hours care and off-hours care, or care classified as "emergency"

---

[3]J-PAL (2014) provides a brief overview of the experiment and some of its findings. More details on the experimental design, as well as specific results can be found in the original papers: Finkelstein et al. (2012), Baicker et al. (2013), Taubman et al. (2014), Baicker et al. (2014), and Finkelstein et al. (2016).

[4]Brot-Goldberg et al. (2017) report qualitatively similar patterns in their (non-randomized) analysis of the effect of the introduction of a high deductible in the context of employer-provided health insurance: it appears to reduce both "high value" and "low value" care similarly.

and care classified as "non emergency"), and is persistent across the two years of the study (Taubman et al., 2014; Finkelstein et al. 2016).

The finding that Medicaid increases use of the emergency department was greeted with considerable attention and surprise (e.g., Beck 2014; Heintzman et al. 2014; Tavernise 2014). Conceptually, however, the result should not be surprising. EMTALA requires hospitals to provide emergency care on credit and prohibits them from delaying treatment to inquire about insurance status or means of payment. Hospitals, however, can – and do – charge the patient for such visits, and Medicaid coverage reduces the out-of-pocket price of the visit to zero, presumably leading to an increase in demand for emergency department visits. At the same time, Medicaid coverage also reduces the price of other care to zero, generating additional, indirect effects, which could go in either direction. Many conjecture that primary care can substitute for emergency department care, and thus cheaper primary care may lead to a reduction in emergency department visits. Yet, the effect could also go in the other direction; for example, one may be more likely to seek emergency room care if one has insurance to cover any recommended follow up treatments. Since the Oregon experiment didn't independently vary the price of primary care and emergency department care, it is not designed to address whether the emergency department and primary care are substitutes or complements. But the results indicate that, on net, Medicaid increases emergency department use, suggesting that any substitution that may exist is not large enough to offset the direct effect of making the emergency department free.

The RAND Health Insurance Experiment. The Oregon Health Insurance Experiment examined the impact of insurance compared to no insurance. A separate question is whether, among those with health insurance, the comprehensiveness of that insurance affects healthcare utilization. Over three decades earlier, in the late 1970s, the RAND Health Insurance Experiment experimentally varied the extent of consumer cost-sharing across about 2,000 non-elderly families in order to study the effect of consumer cost-sharing in health insurance on healthcare spending and health. As before, we focus on the results for healthcare spending only.[5]

---

[5]Our discussion draws heavily on the overview and retrospective provided by Aron-Dine et al. (2013). For more detail on the experimental design and results, readers should consult Newhouse (1993) and the many original research papers discussed and cited therein.

Unlike the Oregon experiment, which was conceived of by policymakers for fairness purposes and capitalized on by academics for research purposes, the RAND Health Insurance Experiment was prospectively designed by researchers to estimate the impact of consumer cost-sharing. Families were randomly assigned to plans for 3-5 years. The plans differed solely in their consumer cost-sharing; for example, one plan had zero cost-sharing ("free plan") while others had 25, 50, or 95 percent cost-sharing (two others set different cost sharing based on the type of care). Importantly, all plans had an out-of-pocket maximum in order to limit the financial exposure of participants; above this maximum amount, families in all plans had full insurance. Thus, referring back to Figure 1, the RAND plans had two of the three coverage arms shown: the coinsurance arm (with coinsurance ranging from zero to 95%), and the catastrophic arm that provides full coverage. The out-of-pocket maximum amounts were set at a fairly low level, so that even the least generous plan had substantial coverage. The exact amount of the out-of-pocket maximum was itself randomly assigned within each co-insurance assignment. The top panel of Figure 3 shows some examples of plans from the RAND experiment. We will return to this aspect of the design in subsequent discussion.

Once again, the results from the randomized evaluation clearly point to the existence of a moral hazard effect. Lower consumer cost-sharing leads to more spending. The bottom panel of Figure 3 provides a flavor of these results, showing how the share of individuals with any annual healthcare spending decreases as the health insurance coverage becomes less generous.

3.2. Quasi-Experimental Evidence: Bunching in Medicare Part D

In addition to the randomized evaluations, a very large number of quasi-experimental studies also show that health insurance coverage is associated with increased healthcare spending. Here we focus on one such example, which is based on prescription drug spending responses to the Medicare Part D prescription drug benefit. It will serve as a subsequent point of departure for the modeling of spending under alternative contracts that is the focus of Section 5 below.

Medicare Part D was launched in 2006 to add prescription drug coverage to the existing

Medicare public health insurance program for the elderly and disabled in the United States. In 2015, Medicare Part D covered about 42 million individuals and generated approximately $77 billion in budgetary outlays (Congressional Budget Office 2015). The original Medicare program – introduced in 1965 to cover hospital and physician services – offers uniform, publicly provided coverage. Medicare Part D, by contrast, is provided by private insurers who are required to offer coverage that is actuarially equivalent or more generous than a government-designed standard benefit.

The top panel of Figure 4 shows the government-defined standard benefit design in 2008. It shows the highly non-linear nature of the standard Part D contract. According to this contract, the individual initially pays for all expenses out of pocket, until she has spent $275 (in cumulative drug spending within the covered year), at which point she pays only 25% of subsequent drug spending until her total drug spending reaches $2,510. At this point the individual enters the famed "donut hole," within which she must once again pay for all expenses out of pocket until total drug spending reaches $5,726, the amount at which catastrophic coverage sets in and the marginal out-of-pocket price of additional spending drops substantially, to about 7%.

As noted, individuals may buy plans that are actuarially equivalent to, or have more coverage than the standard plan, so that the exact contract design varies across individuals. However, a common feature of these plans is the existence of substantial non-linearities that are similar to the standard coverage we have just described. In particular, the location of the "donut hole" at the government-set kink location is typical of most plans, although some of these plans provide partial coverage within the donut hole region. Using data on Medicare Part D beneficiaries from 2007-2009, we estimated that a beneficiary entering the coverage gap experiences, on average, a price increase of almost 60 cents for every dollar of total spending (Einav, Finkelstein and Schrimpf 2015).

As many economists have observed, the donut hole is incompatible with basic economic theory, which would imply greater coverage for greater financial loss, or a concave coverage function as in Figure 1. The donut hole apparently arose as a political compromise between the objective of having a program in which even those who spend little on drugs receive benefits and the need to keep projected expenditures below the legislated cap (Duggan,

Healy and Scott Morton 2008).

Whatever its theoretical demerits or political origins, the donut hole has proved a boon for empirical research on the moral hazard effects of insurance. Standard economic theory suggests that, as long as preferences for healthcare and consumption are strictly convex and smoothly distributed in the population, we should expect the distribution of individuals' spending to bunch at a convex kink point of their budget set. This suggests a natural test for a behavioral response to price. If moral hazard does not exist, individual spending will be distributed smoothly in the population. With moral hazard, bunching will be observed around the convex kink in the budget set at the start of the donut hole, where insurance becomes discontinuously less generous on the margin.[6] Indeed, the bottom panel of Figure 4 shows a histogram of total annual prescription drug spending in 2008. The response to the convex kink at the donut hole is apparent: there appears to be a noticeable spike in the distribution of annual spending around the kink location. Moreover, the government changes the kink location each year and the location of the bunching moves in virtual lock step as the location of the kink moves. Across all years, we estimate that the convex kink leads to a statistically significant 29% increase in the density of individuals whose annual spending is around the kink location (Einav, Finkelstein and Schrimpf 2015).

4. The Nature of Moral Hazard in Health Insurance

4.1. What is "The Price" of Medical Care in The Presence of Non-Linear Contracts?

We view the results summarized in the last section as presenting compelling evidence that moral hazard in health insurance exists: healthcare spending is higher when insurance coverage increases. This evidence seems a natural and necessary pre-condition for spending time and effort to model what spending would be under alternative contracts. This is one

---

[6]This idea that individuals will bunch at convex kinks in their budget set has been present in the literature since the late 1970s. In the last decade, the increased availability of large and detailed administrative data has helped spur an explosion of empirical work on bunching, initially in the context of labor supply responses to the non-linear income tax schedule (e.g. Saez, 2010), but also in other contexts. Kleven (2016) provides an excellent review of this growing literature.

– presumably simple and obvious but important nonetheless – way in which reduced form work can complement economic modeling.

Yet, the evidence we have shown thus far provides little guidance regarding the nature of this moral hazard response or, relatedly, regarding the appropriate economic model to apply to the data. The non-linear nature of virtually all health insurance contracts in the United States raises a key modeling question: what is the price of healthcare perceived by the insured individual as she contemplates using a specific healthcare service? Put differently, to what extent do individuals respond to the dynamic incentives that are generated by the non-linearity of the health insurance coverage?

Until recently, this question had attracted relatively little attention in the moral hazard literature. Instead, a large number of empirical studies endeavored to summarize the impact of health insurance on healthcare utilization by reporting the price elasticity of the demand for medical care with respect to "the" out-of-pocket price. A review article by Cutler and Zeckhauser (2000), for example, summarizes about 30 such studies. A particularly famous and widely-used estimate is the RAND Health Insurance Experiment's estimate of the price elasticity of demand for medical care of -0.2 (Manning et al. 1987; Keeler and Rolph 1988).

However, in the presence of non-linear contracts, applying such single elasticity estimates is challenging without some guidance as to whether and how one can map a non-linear insurance coverage into a single price. For example, one cannot extrapolate from estimates of the effect of co-insurance on healthcare spending to the effects of introducing a high-deductible health insurance plan without knowing how forward looking individuals are in their response to health insurance coverage and their beliefs about the distribution of future health shocks. A completely myopic individual would respond to the introduction of a deductible as if the price has sharply increased to 100%, whereas a fully forward looking individual with annual healthcare spending that are likely to exceed the new deductible would experience little change in the effective marginal price of care.

The original RAND investigators were, of course, acutely aware of this issue and spent considerable effort estimating and modeling how individuals respond to the non-linear incentives in the RAND contracts (Keeler and Rolph 1988). However, application of their -0.2 estimate in a manner consistent with their model is a non-trivial manner. Although

14

notable exceptions exist (e.g. Buchanan et al. 1991; Keeler et al. 1996), most subsequent researchers have applied the RAND estimates in a much simpler fashion: they summarized the non-linear insurance contracts with a single price to which the -0.2 elasticity was applied. For example, researchers used the average out-of-pocket price (Newhouse 1992; Cutler 1995; Cogan, Hubbard and Kessler 2005; Finkelstein 2007), the realized end-of-year price (Eichner 1998; Kowalski 2016), or the expected end-of-year price (Eichner 1997) as various ways to summarize the non-linear contract with a single number.

These choices can be consequential for the magnitude of the predicted spending response. Consider for example an attempt to forecast the effect of changing the plan from the RAND plan with a 25% percent coinsurance plan (and its associated, randomly assigned out-of-pocket maximums) to a plan with a constant 28% coinsurance plan. The price of medical care under the constant 28% coinsurance plan is well-defined (0.28). But in order to directly apply the RAND estimate of -0.2, we would also need to summarize the non-linear RAND plan with a 25% coinsurance and a given out of pocket maximum with a single price; this essentially means choosing the weights to construct an average price. In Aron-Dine et al. (2013) we showed that three different ways to map the non-linear RAND contract to a single price lead to out-of-sample spending predictions for the 28% constant co-insurance contract that vary by a factor of 2.

This shows that more work and care is needed to thoughtfully apply out-of-sample the results from even a justifiably famous and well-designed randomized experiment. While the RAND health insurance experiment was prospectively designed to analyze the impact of cost sharing, at the end what it delivers is estimates of the causal effect of specific (non-linear) health insurance plans. In order to move beyond what the experiment directly delivers - estimates of specific plans' "treatment effects" - more assumptions regarding an economic model of behavior are needed. The RAND estimates continue to be used to this day in forecasting the effects of actual and proposed policies. Given the hard work that went into deriving those credible reduced form estimates, it seems hard to argue with devoting a commensurate amount of effort to considering how one might sensibly transform them out of sample.

4.2. Do Individuals Respond to Dynamic Incentives?

Once we recognize that the treatment of the non-linear budget set can be consequential for this out-of-sample translation, the first question is whether in fact individuals take the dynamic incentives that are associated with the non-linear budget set into account. A fully rational, forward-looking individual who is not liquidity constrained should take into account only the future price of medical care and recognize that (conditional on that future price) the current spot price on care is not relevant, and should not affect healthcare utilization decisions. However, there are a number of reasons why individuals might respond only to the spot price. They may be (or behave as if they are) unaware of or not understand the non-linear budget set created by their health insurance contract, they may be affected by an extreme form of present bias and behave as if they are completely myopic, or they may wish to factor in the future price but be affected entirely by the spot price due to liquidity constraints.

The ideal way to test the null hypothesis of whether dynamic incentives matter would be to hold the spot price of care constant while varying the future price of care. As it turns out, the RAND Health Insurance Experiment did exactly that! As mentioned in Section 3 (see Figure 3), the RAND experiment randomly assigned the co-insurance rate across families and then, within each coinsurance rate, randomly assigned families to different levels of the out-of-pocket maximum. In principle, this is precisely the variation needed to test the null of whether individuals respond to the dynamic incentives: one would want to compare the initial healthcare utilization decisions of individuals randomized into plans with the same coinsurance rate but different out-of-pocket maximum. In practice, however, this approach is hampered by the relatively small sample sizes in the RAND experiment as well as the relatively low levels of the plans' maximum amounts (Aron-Dine et al. 2015).

In the absence of the ideal experimental variation, in Aron-Dine et al. (2015) we instead take advantage of a particular feature of many U.S. health insurance contracts that generates quasi-experimental variation that is conceptually similar to this ideal. Most health insurance contracts are annual and reset on January 1, regardless of when coverage began. When individuals join a plan in the middle of the year, the deductible and other cost sharing features

remain at the annual level, but are applied for a shorter coverage period. As a result, people who join the same plan in different months of the year face different contract lengths and therefore potentially different future prices, even though they all begin with the same spot price. A test of whether individuals respond to dynamic incentives then becomes whether individuals who join the same plan in different months of the year – and therefore face the same initial spot price of care but different future prices – have different initial healthcare utilization. We applied this idea in two settings: employer-provided health insurance and Medicare Part D. In both settings we were able to reject the null that individuals respond only to the spot price of care: individuals who faced the same spot price but higher future prices used less healthcare initially.

Figure 5 summarizes the nature of our findings in the Medicare Part D context. Medicare Part D annual plan choices are typically made during the open enrollment period in November and December, and provide coverage from January to December of the following year. However, when individuals become newly eligible for Part D at age 65, they can enroll in a plan the month they turn 65; the plan's cost-sharing features reset on January 1, regardless of when in the year the individual enrolled. Variation in birth month thus generates variation in contract duration, and hence potentially in expected end-of-year price among individuals in a given plan in their first year.

Figure 5 shows future prices and initial claims for 65 year olds who enrolled in Medicare Part D between February and October. It shows the pattern of future prices and initial claims by enrollment month, separately for beneficiaries in two groups of plans: deductible and no-deductible plans (recall that the standard benefit design has a deductible, but insurers can offer more generous coverage than the standard design; many offer no-deductible options). We measure initial drug use by whether the individual had a prescription drug claim in the first three months of coverage. We summarize the dynamic incentives in the contract with the expected end-of-year price. The expected end-of-year price depends on three elements: the cost-sharing features of the beneficiary's plan, the duration (number of months) of the contract (which in turn is determined by their birth month), and the beneficiary's expected spending (which we calculate based on the spending of all individuals who enrolled in that plan in that month). Of course, if individuals do not believe their spending risk is drawn

17

from the same distribution as everyone else who joined their plan in their month, there will be measurement error in the expected end-of-year price; similarly, if individuals are not risk neutral, other moments of the distribution of the end-of-year price may affect their initial utilization. Such modeling choices could be consequential if our goal were to estimate the extent of forward looking behavior. They may also bias us against rejecting the null of no forward looking behavior. However, if we do reject that null despite such potential sources of measurement error, it is informative.

The results provide evidence against the null that individuals do not respond to the future price. In the deductible plan, Figure 5 shows that the expected end-of-year price is increasing in the enrollment month; a later enrollment date gives the individual less time to spend past the deductible and into the lower consumer cost-sharing arm. Recall that all individuals in these plans face the same initial spot price of care; what varies is the contract length and thus the expected end-of-year price. In these plans, we see that initial utilization is decreasing with enrollment month. By contrast, in the no-deductible plan, the expected end-of-year price is decreasing with the enrollment month; here, a later enrollment date gives the individual less time to spend past the cost-sharing arm and into the donut hole. In these plans, by contrast, the probability of an initial claim does not appear to vary systematically with the enrollment month. Combined, the contrast suggests that, holding the spot price of care constant, initial healthcare use is decreasing in the expected end-of-year price. In other words, individuals appear to respond to the dynamic incentives.

5. Forecasting Healthcare Spending under Counterfactual Contracts

The descriptive results from the last two sections suggest that individuals' decision making regarding healthcare utilization responds to the insurance coverage, and that this response is affected by the dynamic incentives associated with the non-linear health insurance contracts commonly offered in the United States. One clear implication of these results is that assuming that the spot price associated with a given medical treatment is the only relevant price is

problematic. However, we cannot conclude from this evidence that consumers do not respond at all to the spot price. Indeed, there is evidence to the contrary: Brot-Goldberg et al. (2017) study the introduction of a high-deductible plan (where previously there was no deductible) and present evidence that suggests a response to the spot price as well: predictably sick consumers reduce their spending in response to the deductible, despite the fact that they are likely to end the year outside of the deductible range. They conclude that changes in the spot price – rather than the future price – are the primary drivers of the reduced spending they observe when the high deductible is introduced.

When individuals respond to both spot and future prices, summarizing a given contract with a single price is not a sensible option. Therefore, when researchers want to use the experimental (or quasi-experimental) results to provide predictions for spending under other, counterfactual contracts not seen in the data, a more complete behavioral model is needed. We undertook such exercises in two related papers (Einav, Finkelstein and Schrimpf 2015; Einav, Finkelstein and Schrimpf 2017). Our goal was to analyze spending under alternative non-linear Part D contracts, and our motivating point of departure was the bunching at the convex kink created by the donut hole, which we described earlier. We showed that two different – and in our subjective opinion "reasonable" – models could both match the observed bunching, but produce fairly different out-of-sample predictions. This underscores the importance of modeling choices in extrapolating out of sample. Ideally, other evidence can be brought to bear to guide model selection.

In our context, we developed two alternative, non-nested models. One natural approach we implement is to adapt the Saez (2010) framework to our context. In this influential paper, Saez (2010) showed how a stylized, static, frictionless model of labor supply can allow for a simple mapping from the observed bunching around convex kinks in the income tax schedule to an estimate of the elasticity of labor supply. In Einav, Finkelstein and Schrimpf (2017) we translated Saez's model of labor supply to a model of prescription drug spending and applied his approach straightforwardly to the Medicare Part D setting. To do so, we assumed that individual i has quasi-linear utility in drug spending (m) and residual income (y): $u_i(m, y) = g_i(m) + y$. We chose a particular functional form for $g_i(m)$ so as to obtain a constant elasticity form for drug spending as a function of the out-of-pocket price that would

be similar to Saez's constant elasticity form for hours of work with respect to the after-tax wage. This allowed us to almost exactly follow his strategy and derive a mapping between the observed extent of bunching around the donut hole and the elasticity of drug spending with respect to the out-of-pocket price. This exercise resulted in an estimated elasticity of drug spending with respect to the out-of-pocket price of about -0.05. Because this is based on the bunching at the kink in annual drug spending, the spot and the future price of care are the same for the "bunchers" at the end of the year, which makes this a well-defined object.

Of course, the simplicity of the Saez-style approach comes at the cost of potentially abstracting from a host of real-world features that may be important in a particular context. Our real-world problem is dynamic: individuals make sequential purchase decisions throughout the year as information is revealed, and they make current healthcare utilization decisions facing uncertain future health shocks. The reduced form evidence we discussed in the previous section suggests that individuals do not ignore the future in making such decisions. This reduced form evidence has implications for model selection. In particular, it suggests that a static model - such as our adaptation of Saez (2010) - may miss some important features of the consumer problem.

We therefore also developed a dynamic model of drug use in which a (potentially) forward looking individual facing uncertain future health shocks makes drug purchase decisions (Einav, Finkelstein, and Schrimpf 2015). We modeled weekly drug spending decisions, where each week there is some chance of a health event that could be treated by a prescription; if it occurs, the individual must decide whether or not to fill the prescription that week. The individual is covered by a non-linear prescription drug insurance contract over 52 weeks. A coverage contract is given by a function, similar to the one depicted in the top panel of Figure 4, that specifies the out-of-pocket amount the individual would be charged for a prescription drug with a given list price given the cumulative out-of-pocket spending up until that point in the coverage period. Optimal behavior can be characterized by a simple finite horizon dynamic problem. The three state variables are the number of weeks left until the end of the coverage period, the total amount spent so far, and a health state, which accounts for potential serial correlation in health.

20

In this model there are three economic objects. The first is a statistical description of the distribution of health shocks. The second key object is the primitive price elasticity, or "moral hazard," that captures contemporaneous substitution between health and income. The third object captures the extent to which individuals understand and respond to the dynamic incentives associated with the non-linear contract. As discussed in the last section, there is evidence that this response exists. The model allows us to quantify it, and to translate it into implications for annual drug spending under alternative - potentially counterfactual - contracts.

We parameterized the model with distributional and functional form assumptions and estimated it using simulated minimum distance. Importantly, one of the moments we fit is the extent of bunching around the donut hole. We then used the estimates to simulate the spending response to a uniform percentage price reduction in all arms of the standard, government-defined plan; this yields implied elasticities of about -0.25. This elasticity estimate is five times higher than what the Saez-style static model produced.

Thus, both the static Saez-style model and the dynamic model match, by design, the same observed bunching pattern, but they deliver very different out-of-sample predictions. The appeal of the Saez-style model is the simple and transparent mapping from the descriptive fact to the economic object of interest; relatedly, it can be implemented relatively quickly and easily. The dynamic model is more computationally challenging and time consuming to implement; it also has (despite our best efforts) more of a "black box" relationship between the underlying data objects and the economic objects of interest. However, it can account for potentially important economic forces that the static model abstracts from. In particular, it can account for anticipatory responses by forward looking agents to changes in the future price. The static model imposes that any response to the donut hole is limited to people around the donut hole. In contrast, the dynamic model allows for the possibility that the set of people near the donut hole – and therefore "at risk" of bunching – may in fact be endogenously affected by the presence of the donut hole; forward-looking individuals, anticipating the increase in price if they experience a series of negative health shocks, are likely to make purchase decisions that decrease their chance of ending up near the donut hole, even if at that point they are far from reaching it. Indeed, when we considered the implications in

the dynamic model of "filling the donut hole" (i.e. providing 25% coinsurance in the donut hole instead of 100% coinsurance as scheduled under the Affordable Care Act to occur by 2020), we estimated that about one-quarter of the resultant spending increase came from "anticipatory" responses by individuals whose annual spending prior to this policy change would have been well below the donut hole (Einav, Finkelstein and Schrimpf 2015).

The comparison of the results from the static and dynamic model highlights a broader point that should be neither novel nor surprising: modeling choices are consequential. In this specific application, we show that an in-sample bunching pattern may be rationalized by different modeling assumptions, and these assumptions can, at least in some contexts, have very different quantitative implications out-of-sample. This issue is not unique to the bunching literature. The phenomenon is more general. For example, our discussion of the results of the RAND Health Insurance Experiment above illustrated that the assumptions made in translating the experimental treatment effects into economic objects that could be applied out of sample were also consequential.

More generally, the bunching literature following Saez (2010) is one specific application of the influential "sufficient statistics" literature popularized by Chetty (2009) – which attempts to use simple models to directly and transparently map reduced form parameters into economic primitives. Our analysis illustrates that two different models can map the same reduced form object into very different out-of-sample predictions. Sufficient statistics, in other words, are sufficient conditional on the model (or set of models). This is an obvious point, made clearly by Chetty (2009), but sometimes forgotten in applications and interpretations.

6. Conclusions

The title (and purpose) of our paper is to discuss both "what we know" and "how we know it." The research on moral hazard effects of health insurance that we described (hopefully) illustrates the claim we made at the outset: "reduced form" and "structural" work have their different strengths and limitations, and are most powerful when used in tandem (within or across papers) to answer a given question or a related set of questions.

The reduced form evidence tells us unambiguously that health insurance increases health

care utilization and spending. Moral hazard, in other words, irrefutably exists. The overwhelming, compelling evidence on this point – from several randomized evaluations as well as countless, well-crafted quasi-experimental studies – should give any informed reader considerable pause when they hear claims to the contrary. Consider the rhetorical debate we started with over whether moral hazard exists (e.g. Gladwell 2005) and if so whether it might be of the opposite sign. These qualitative hypotheses are powerfully rejected by the reduced form evidence. This is a particular illustration of a broader point: when the debate is about sharp nulls, or qualitative signs, credible reduced form studies, which often rely on fewer modeling assumptions, are very powerful in convincingly distinguishing between competing hypotheses.

Reduced form evidence can also be valuable for retrospective analysis when an existing policy of interest is captured by the reduced form variation. If one is interested in the question: what happened when Oregon expanded Medicaid coverage in 2008, there is no better way to answer that than with the results of the lotteried expansion. Likewise, historical interest in the impact of the original introduction of Medicare can be well-served by reduced form analyses of the impact of that introduction (Finkelstein 2007; Finkelstein and McKnight 2008).

One might also be tempted to use reduced form results for prospective analyses of policies that are "close enough" to the reduced form variation. Here, however, it becomes challenging without additional theory and evidence to know what dimensions of the setting are important and how to judge "closeness" in those dimensions. For example, the low-income, able boded uninsured adults covered by Medicaid through the 2008 Oregon Health Insurance Experiment are a very similar population to the low-income able boded uninsured adults covered by the 2014 Medicaid expansions under the Affordable Care Act; indeed, the only obvious difference is that in Oregon eligibility required the individual to be below 100% of the federal poverty line while the state Medicaid expansions reached to 138% of the federal poverty line. Yet a host of factors could produce differential short-run impacts of Medicaid in Oregon and in these other expansions. The most obvious is that the demographics of low income adults and the nature of the healthcare system (including the healthcare safety net) differs across the country. One could perhaps shed some light on this (power permitting) through

heterogeneity analysis in the Oregon experiment across types of people and places. Other observable differences – such as in the macro economy – would be harder to address. More subtle conceptual differences would require more thought and modeling. For example, the partial equilibrium impacts of covering a small number of people in Oregon might differ from the general equilibrium effects of a market-wide expansion in insurance coverage under the ACA (Finkelstein 2007). The impact of health insurance for individuals who voluntarily sign up for the lottery may well be different than the impact when, as in the ACA, insurance coverage is mandatory (Finkelstein et al. 2012; Einav et al. 2013).

The limitations of prospective policy analysis with reduced form evidence points to the need for economic modeling. More broadly, whenever we want to study the impact of something not observed in the data, we need a model to extrapolate from reduced form estimates to the setting of interest. The results from the RAND Health Insurance Experiment that we described illustrated this point. The RAND experiment delivers causal estimates of the spending impact of the particular health insurance contracts included in the experiment. The literature has since extrapolated from these plan fixed effects to forecast the spending effects of alternative contracts not observed in the data, such as high-deductible plans. As we have seen, the modeling choices made in such extrapolations are quite consequential for the translation of the reduced form estimates into spending forecasts. Since ad hoc choices of how to extrapolate from reduced form estimates to contracts not observed in the data can yield very different results, this suggests the value of more formal modeling in which one specifies and estimates a model of primitives that govern how an individual's medical care utilization responds to the entire non-linear budget set contracted by the health insurance contract.

This is a non-trivial exercise. It requires, among other things, estimating the individual's beliefs about the arrival rate of medical shocks over the year, her discount rate of future events, and her willingness to trade off health and medical utilization against other consumption. Naturally, as we illustrated, the modeling choices themselves will be consequential, even when they can match the reduced form facts. Here, the reduced form evidence that individuals are at least partly forward looking can motivate the use of a dynamic model.

We thus see great complementarity between the reduced form analysis and economic

modeling in ways that our examples have hopefully illustrated. Economic models allow us to get more bang for our reduced form buck – analyzing, for example, not just whether the current Part D contracts affect drug spending but forecasting what that spending would be like under alternative policies. In turn, reduced form evidence allows us to focus our questions – it's useful to verify that moral hazard exists before trying to model it – and make more informed modeling choices.

It's a great time to be an empirical economist. We have a rich tradition of economic modeling and structural estimation to draw upon. And we are the beneficiaries of an improved (and improving!) reduced form tool kit for identifying causal effects (Angrist and Pischke 2010). Both can be applied to the large, and rich administrative data sets that researchers are increasingly accessing. By combining these approaches – within and across papers – our production possibility frontier will expand even further.

References

Akerlof, George (1970). "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics*, 84(3), 488–500.

Angrist, Joshua D., and Jorn-Steffen Pischke (2010). "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics." *Journal of Economic Perspectives*, 24(2), 3-30.

Aron-Dine, Aviva, Liran Einav, and Amy Finkelstein (2013). "The RAND Health Insurance Experiment, Three Decades Later." *Journal of Economic Perspectives*, 27(1), 197-222.

Aron-Dine, Aviva, Liran Einav, Amy Finkelstein, and Mark Cullen (2015). "Moral Hazard in Health Insurance: Do Dynamic Incentives Matter?" *Review of Economics and Statistics*, 97(4), 725-741.

Arrow, Kenneth J. (1963). "Uncertainty and The Welfare Economics of Medical Care." *American Economic Review*, 53(5), 941-973.

Baicker, Katherine, Sarah Taubman, Heidi Allen, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Eric C. Schneider, Bill Wright, Alan M. Zaslabsky, and Amy Finkelstein (2013). "The Oregon experiment - effects of Medicaid on clinical outcomes." *New England Journal of Medicine*, 386(18), 1713-1722.

Baicker, Katherine, Amy Finkelstein, Jae Song, and Sarah Taubman (2014). "The Impact of Medicaid on Labor Market Activity and Program Participation: Evidence from the Oregon Health Insurance Experiment." *American Economic Review Papers and Proceedings*, 104(5), 322-328.

Beck M (2014). "Medicaid expansion drives up visits to ER." *Wall Street Journal*, January 3.

Brot-Goldberg, Zarek C., Amitabh Chandra, Benjamin R. Handel, and Jonathan T. Kolstad (2017). "What Does a Deductible Do? The Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics." *Quarterly Journal of Economics,* 132(3), 1261-1318.

Buchanan, Joan L., Emmett B. Keeler, John E. Rolph, and Martin R. Holmer (1991). "Simulating Health Expenditures under Alternative Insurance Plans." *Management Science*, 37(9): 1067-1090.

Clemens, Jeffrey, and Joshua D. Gottlieb (2014). "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?" *American Economic Review*, 104(4), 1320-1349.

Cogan, John F., R. Glenn Hubbard, and Daniel P. Kessler (2005). *Healthy, Wealthy, and Wise: Five Steps to a Better Healthcare System.* 1st ed. Washington, DC: AEI Press.

Chetty, Raj (2009). "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods." *Annual Review of Economics*, 1, 451-488.

Cutler, David M. (1995). "Technology, Health Costs, and the NIH." Paper prepared for the National Institutes of Health Economics Roundtable on Biomedical Research. http:// www.economics.harvard.edu/files/faculty/13 _Technology,%20Health%20Costs%20and%20 the%20NIH.pdf.

Cutler, David M., and Richard J. Zeckhauser (2000). "The Anatomy of Health Insurance." In *Handbook of Health Economics*, edited by A. J. Culyer and J. P. Newhouse, volume 1, 563-643. Amsterdam: Elsevier.

Duggan, Mark, Patrick Healy, and Fiona Scott Morton (2008). "Providing Prescription Drug Coverage to the Elderly: America's Experiment with Medicare Part D." *Journal of Economic Perspectives*, 22(4), 69-92.

Dudiak, Z. "Pittsburgh area legislators react to governor's budget proposals, " (2013); http://foresthills-regentsquare.patch.com/groups/politics-and-elections/p/pittsburgh-area-legislators-react-to-governor-s-budge5c772c0e4b.

Ehrlich, Isaac, and Gary S. Becker (1972). "Market insurance, self-insurance, and self-protection." *Journal of Political Economy*, 80(4), 623-648.

Eichner, Matthew J. (1997). "Medical Expenditures and Major Risk Health Insurance." PhD Dissertation, Chapter 1, MIT.

Eichner, Matthew J. (1998). "The Demand for Medical Care: What People Pay Does Matter." *American Economic Review Papers and Proceedings*, 88(2), 117-121.

Einav, Liran, and Amy Finkelstein (2011). "Selection in Insurance Markets: Theory and Evidence in Pictures." *Journal of Economic Perspectives* 25(1): 115-138.

Einav, Liran, Amy Finkelstein, Steven P. Ryan, Paul Schrimpf, and Mark R. Cullen (2013). "Selection on moral hazard in health insurance." *American Economic Review*, 103(1), 178-219.

Einav, Liran, Amy Finkelstein, and Paul Schrimpf (2015). "The Response of Drug Expenditure to Nonlinear Contract Design: Evidence from Medicare Part D." *Quarterly Journal of Economics*, 130(2), 841-899.

Einav, Liran, Amy Finkelstein, and Paul Schrimpf (2017). "Bunching at the kink: implications for spending responses to health insurance contracts." *Journal of Public Economics*, 146, 27-40.

Einav, Liran, Amy Finkelstein, and Neale Mahoney (2017). "Provider Incentives and Healthcare Costs: Evidence from Long-Term Care Hospitals." NBER Working Paper No. 23100.

Einav, Liran, and Jonathan Levin (2010). "Empirical Industrial Organization: A Progress Report." *Journal of Economic Perspectives*, 24(2), 145-162.

Eliason, Paul J., Paul L. E. Grieco, Ryan C. McDevitt, and James W. Roberts (2016). "Strategic Patient Discharge: The Case of Long-Term Care Hospitals." NBER Working Paper No. 22598.

Feldstein, Martin (1973). "The Welfare Loss of Excess Health Insurance." *Journal of Political Economy*, 81(2), 251-280.

Finkelstein, Amy (2007). "The Aggregate Effects of Health Insurance: Evidence from the Introduction of Medicare." *Quarterly Journal of Economics*, 122(1), 1-37.

Finkelstein, Amy, and Robin McKnight (2008). "What did Medicare do? The initial impact of Medicare on mortality and out of pocket spending." *Journal of Public Economics*, 92(7), 1644-1668.

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph E. Newhouse, Heidi Allen, and Katherine Baicker (2012). "The Oregon Health Insurance Experiment: evidence from the first year." *Quarterly Journal of Economics*, 127(3), 1057-1106.

Finkelstein, Amy N., Sarah L. Taubman, Heidi L. Allen, Bill J. Wright, and Katherine Baicker (2016). "Effect of Medicaid Coverage on ED Use - Further Evidence from Oregon's Experiment." *New England Journal of Medicine*, 375(16), 1505-1507.

Gladwell, Malcolm (2005). "The Moral-Hazard Myth." *New Yorker*, August 29.

Heckman James J. (2010). "Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy." *Journal of Economic Literature*, 48(2), 356-398.

Heintzman John, Rachel Gold, Steffani R. Bailey, and Jennifer E. DeVoe (2014). The Oregon experiment re-examined: the need to bolster primary care. *BMJ*, 349.

Holmstrom, Bengt (1979). "Moral Hazard and Observability." *Bell Journal of Economics*, 10(1), 74-91.

Imbens, Guido W. , and Joshua D. Angrist (1994). "Identification and Estimation of Local Average Treatment Effects" *Econometrica*, 62(2), 467-475.

J-PAL (2014). "Insuring the Uninsured." https://www.povertyactionlab.org/sites/default/files/publications/Insuring_the_Uninsured.pdf.

Keeler, Emmett B., and John E. Rolph (1988). "The Demand for Episodes of Treatment in the Health Insurance Experiment." *Journal of Health Economics*, 7(4), 337-367.

Keeler, Emmet B., Jesse D. Malkin, Dana P. Goldman, and Joan L. Buchanan (1996). "Can Medical Savings Accounts for the Nonelderly Reduce Health Care Costs?" " *Jounral of American Medical Association,* 275(21), 1666-1671.

King, Gary, Emmanuela Gakidou, Kosuke Imai, Jason Lakin, Ryan T. Moore, Clayton Nall, Nirmala Ravishankar, Manett Vargas, Martha María Téllez-Rojo, Juan Eugenio Hernández-Ávila, Mauricio Hernández-Ávila, and Hector Hernández Llamas (2009). "Public Policy for the Poor? A randomised assessment of the Mexican universal health insurance programme," *The Lancet*, 9673(373), 1447-1454.

Kleven, Henrik (2016). "Bunching." *Annual Review of Economics*, 8, 435-464.

Kowalski, Amanda E. (2016). "Censored Quantile Instrumental Variable Estimates of the Price Elasticity of Expenditure on Medical Care." *Journal of Business and Economic Statistics*, 34(1), 107-117.

Manning, Willard G., Joseph P. Newhouse, Naihua Duan, Emmett B. Keeler, and Arleen Leibowitz(1987). "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment." *American Economic Review*, 77(3), 251–77.

Nevo, Aviv, and Michael D. Whinston (2010). "Taking the Dogma Out of Econometrics: Structural Modeling and Credible Inference." *Journal of Economic Perspectives*, 24(2), 69-82.

Newhouse, Joseph P (1992). "Medical Care Costs: How Much Welfare Loss?" *Journal of Economic Perspectives*, 6(3), 3-21.

Newhouse, Joseph P. (1993). *Free for All? Lessons from the RAND Health Insurance Experiment.* Harvard University Press: Cambridge, MA.

Palm-Houser, S. "Governor Kasich includes Medicaid expansion in proposed Ohio budget" (2013); www.examiner.com/article/governor-kasich-includes-medicaid-expansion-proposed-ohio-budget.

Pauly, Mark V. (1968). "The Economics of Moral Hazard: Comment." *American Economic Review*, 58(3), 531-537.

Rothschild, Michael, and Joseph E. Stiglitz (1976). "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *Quarterly Journal of Economics*, 90(4), 629-649.

Saez, Emmanuel (2010). "Do Taxpayers Bunch at Kink Points?" *American Economic Journals: Economic Policy*, 2(3), 180-212.

Snyder, R. "Facts about Medicaid expansion: Improving care, saving money" (2013); www.michigan.gov/documents/snyder/Medicaid_expansion_-_factsheet_ final_2-6-13_410658_7.pdf.

Spenkuch, Jorg L. (2012). "Moral Hazard and Selection among the Poor: Evidence from a Randomized Experiment." *Journal of Health Economics*, 31, 72-85.

Taubman, Sarah L., Heidi L. Allen, Bill J. Wright, Katherine Baicker, and Amy N. Finkelstein (2014). "Medicaid Increases Emergency Department Use: Evidence from Oregon's Health Insurance Experiment." *Science*, 343(6168), 263-268.

Tavernise S. (2014) "Emergency visits seen increasing with health law." *New York Times*, January 2.

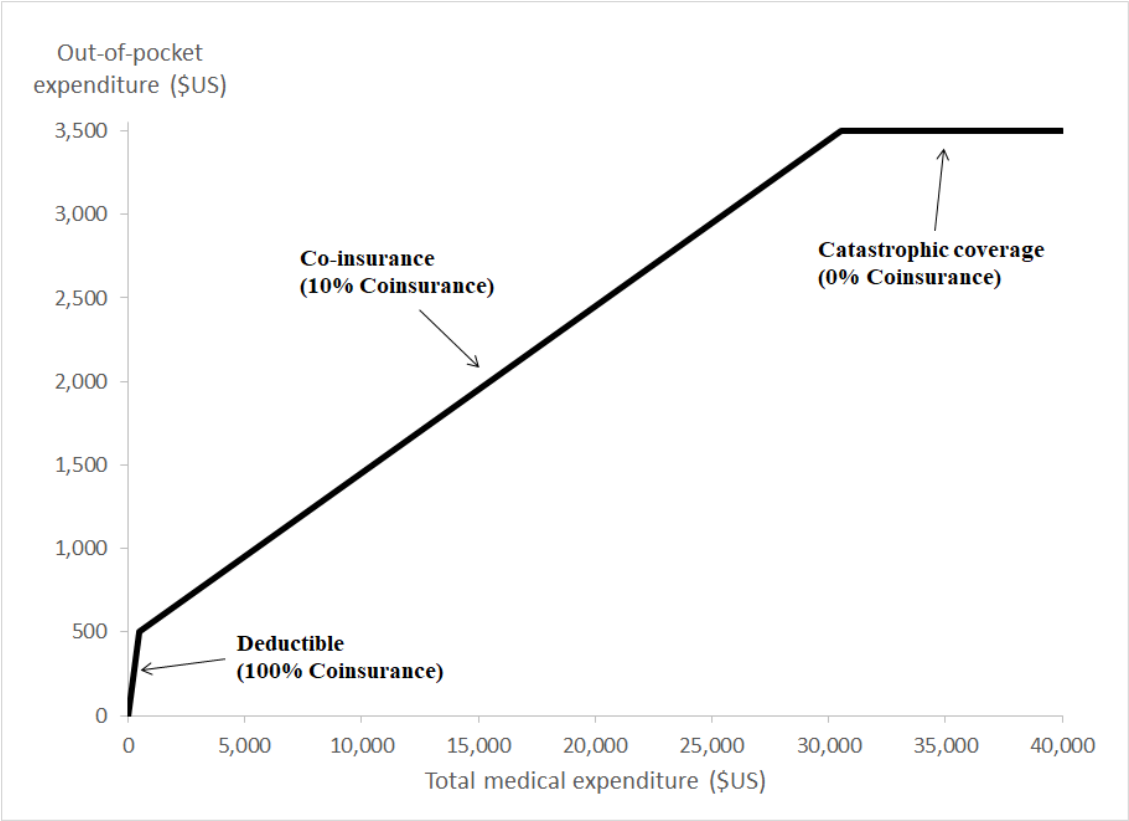## Figure 1: A Typical Health Insurance Contract in the United States

Figure shows a stylized annual health insurance contract, illustrating the mapping the contract creates from total medical spending to out of pocket medical spending. The x-axis shows total medical spending for the year and the y-axis shows the out-of-pocket medical spending for the year.

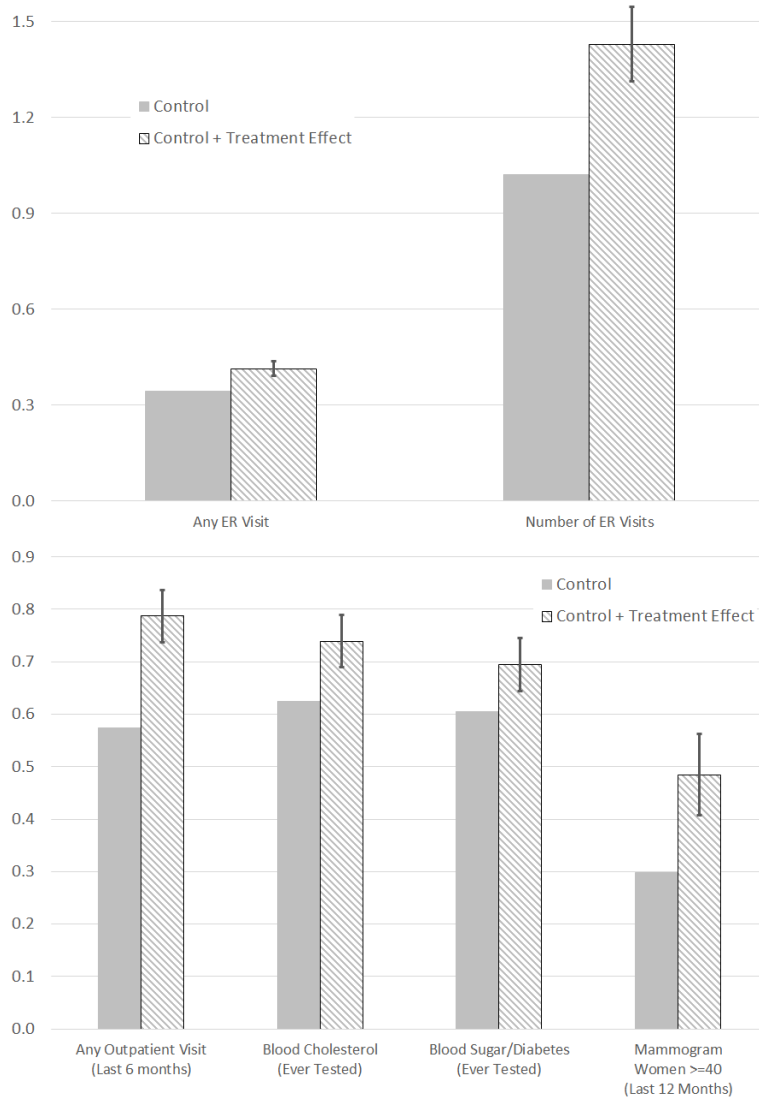Figure 2: Selective Results from The Oregon Experiment

Figure shows selected results from the Oregon Health Insurance Experiment. "Control mean " shows mean for lottery participants who were not selected. "Treatment effect" represents the IV estimate of the impact of Medicaid, using selection by the lottery as an instrument for Medicaid coverage (the first stage is about 0.25). 95 percent confidence intervals are shown with the whisker plot. Top panel shows results for Emergency Room use based on administrative data in the 18 months following the lottery (source: Taubman et al., 2014). Bottom panel shows results for primary and preventive care based on a mail survey administered approximately one year after the lottery (Finkelstein et al. 2012).

Figure 3: Contracts and Outcomes in The RAND Experiment
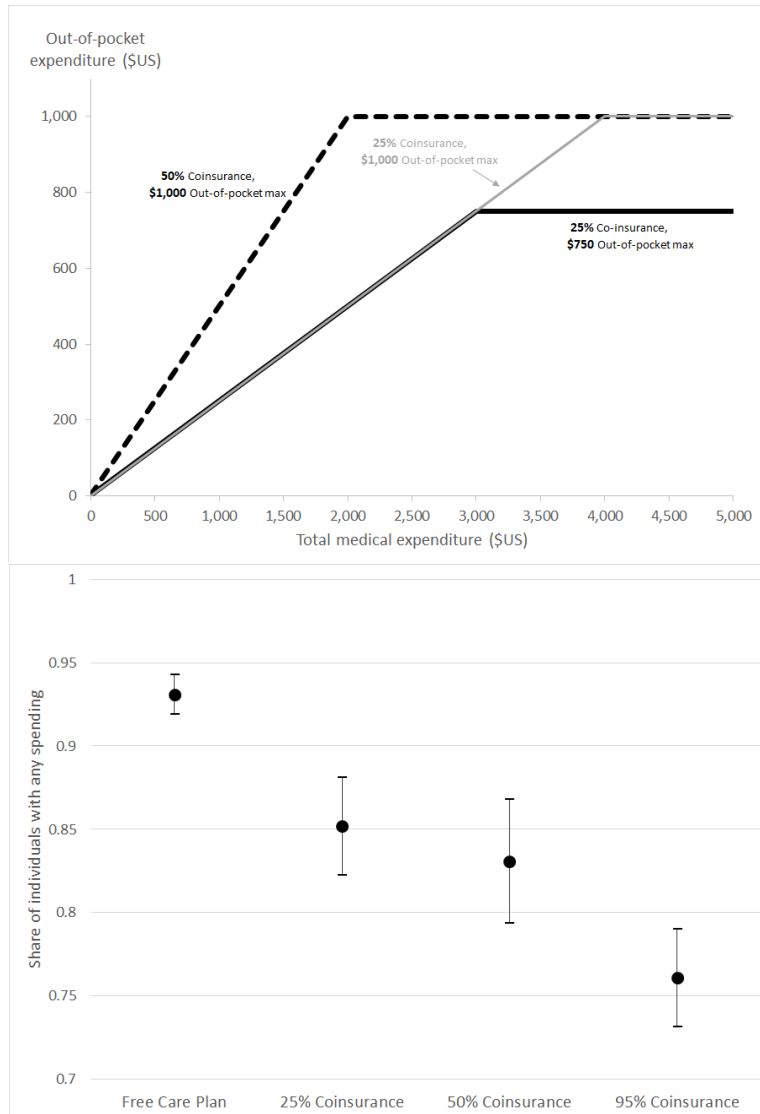
Top panel shows several of the contracts that were randomly assigned to different families in the RAND health insurance experiment; these contracts vary both in their co-insurance and (within coinsurance rates) in their out-of-pocket maximum. Bottom panel reports the estimated treatment effects of the different plans (defined by their coinsurance rate) on the probability of the individual having any medical spending in the year. Source (Aron-Dine et al., 2013, Table 2; see notes therein for more details).

Figure 4: Contract Design and Bunching in Medicare Part D
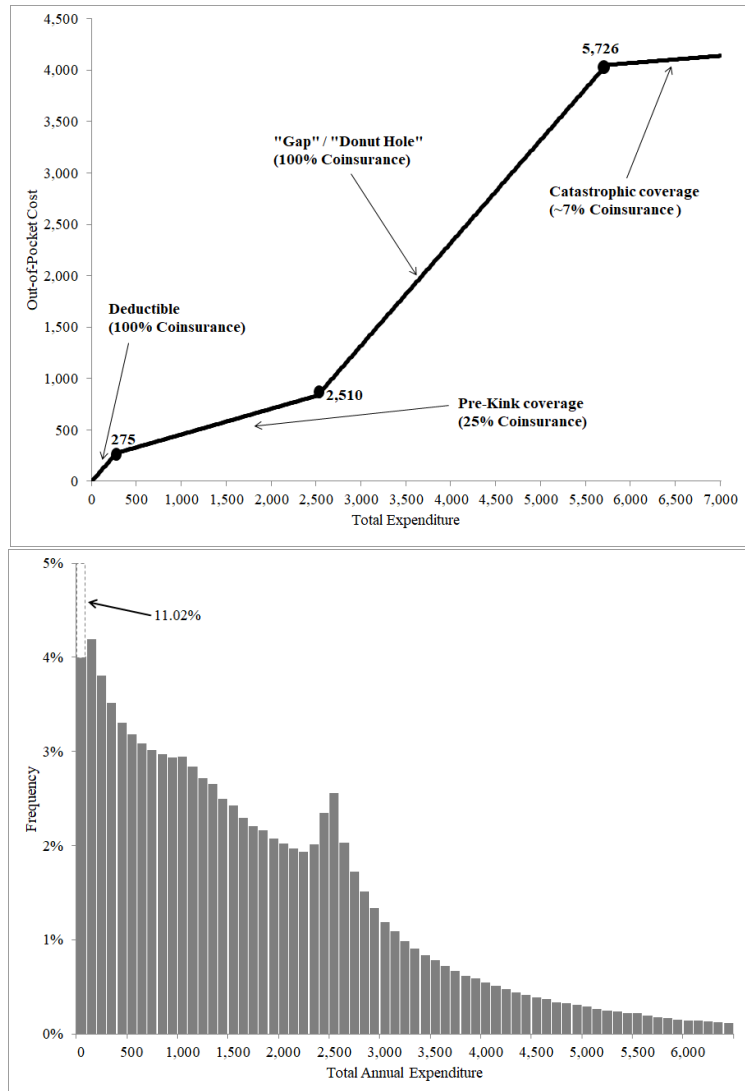
This figure replicates Figure I and Figure II in Einav, Finkelstein, and Schrimpf (2015). Top panel shows the standard benefit design in 2008. "Pre-Kink coverage" refers to coverage prior to the Initial Coverage Limit (ICL) which is where there is a kink in the budget set and the gap, or donut hole, begins. As described in the text, the actual level at which the catastrophic coverage kicks in is defined in terms of out-of-pocket spending (of $4,050), which we convert to the total expenditure amount provided in the figure. Once catastrophic coverage kicks in, the actual standard coverage specifies a set of co-pays (dollar amounts) for particular types of drugs, while in the figure we use instead a 7% co-insurance rate, which is the empirical average of these co-pays in our data. Bottom panel displays the distribution of total annual prescription drug spending in 2008 for our baseline sample. Each bar represents the set of people that spent up to $100 above the value that is on the x-axis, so that the first bar represents individuals who spent less than $100 during the year, the second bar represents $100-200 spending, and so on. For visual clarity, we omit from the graph the 3% of the sample whose spending exceeds $6,500. The kink location (in 2008) is at $2,510. N =1,251,984.

Figure 5: Initial Healthcare Utilization and Future Price
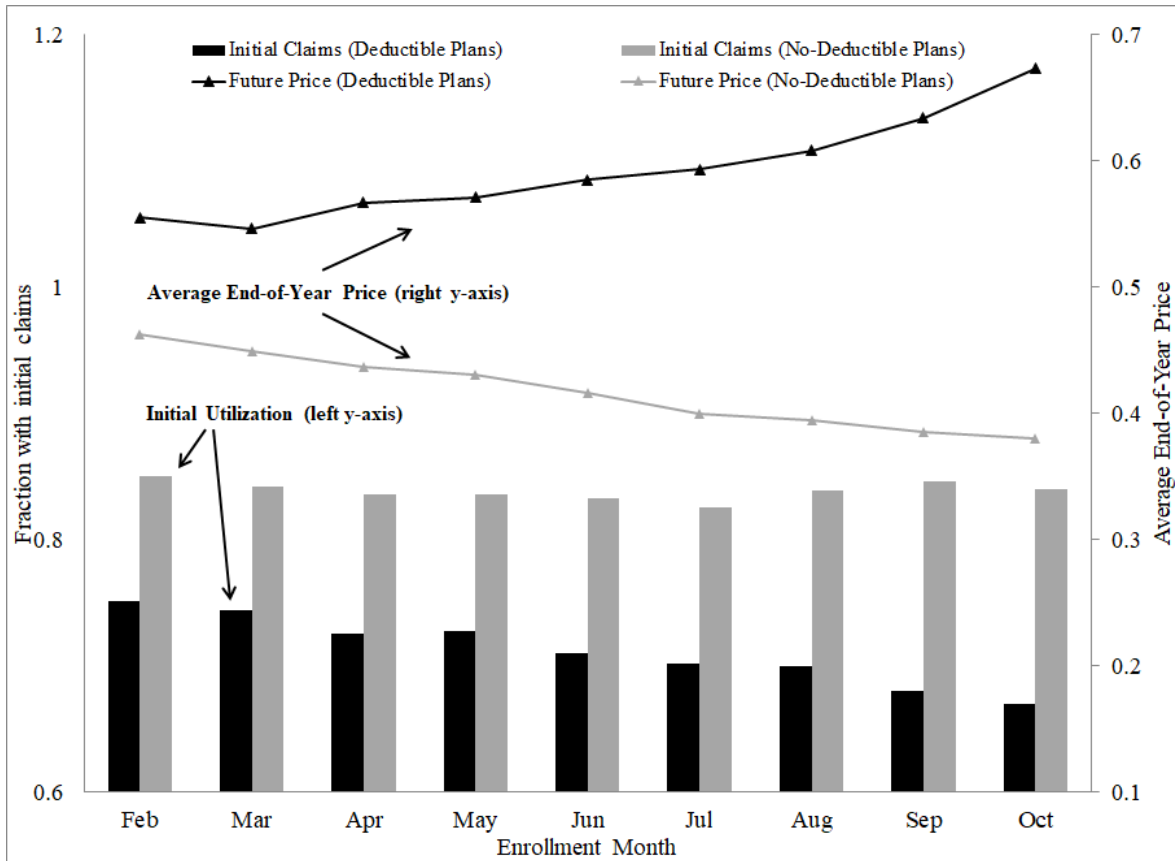
This figure replicates Figure 2 in Aron-Dine et al. (2015). It graphs the pattern of expected end-of-year price and of any initial drug claim by enrollment month for individuals in Medicare Part D during their first year of eligibility (once they turn 65). We graph results separately for individuals in deductible plans and no deductible plans. We calculate the expected end-of-year price separately for each individual based on his plan and birth month, using all other individuals who enrolled in the same plan that month. The fraction with initial claim is measured as the share of individuals (by plan type and enrollment month) who had at least one claim over the first three months. N =137,536 (N=108,577 for no deductible plans, and N=28,959 for deductible plans).